基于互联网的工业设计云数据统计分析研究

景广超，姜立军，邓学雄，李哲林，杨强
（华南理工大学，广州 510006）

摘要：目的 研究工业设计云创新的发展现状和未来趋势。**方法** 提出假设，假定与云创新密切相关的若干关键指标，即广义设计师与设计主题之间的交互数量与交互频率近年来呈现上升态势。基于超图理论和超网络模型，借助模型中的关联矩阵，提出用以表征工业设计云创新发展状况的定量化表征函数，作为统计分析的统计量。基于数据挖掘的思想，使用网络爬虫工具对互联网虚拟设计社区进行数据采集，并结构化存储在关系型数据库中。利用SQL等计算机编程语言对结构化的数据进行处理和统计量计算，绘制出表征函数与年份的XY散点图，并对表征函数与年份作相关分析，根据分析结果来验证假设成立。**结论** 指出工业设计云创新近年来的发展优势和不足，并提出开发一种开放式工业设计云网络平台的建议。

关键词： 工业设计；云创新；数据挖掘；虚拟设计社区；超网络模型；超图；假设检验；相关分析
中图分类号： TB472 **文献标识码：** A **文章编号：** 1001-3563(2014)22-0058-05

Statistical Analysis of Cloud Data for Industry Design Based on Internet

JING Guang-chao, *JIANG Li-jun*, *DENG Xue-xiong*, *LI Zhe-lin*, *YANG Qiang*
(South China University of Technology, Guangzhou 510006, China)

ABSTRACT: Objective To study the development status and future trend of cloud innovation for industry design. **Methods** As key indicators for cloud innovation, the interaction number and interaction frequency between general designers and design themes are hypothesized to be ascending. Based on the theory of hyper-graph and incidence matrix in hyper-network model, quantitative characterization functions are formulated as statistics for the development of cloud innovation.Based on the ideology of data mining, the data is captured from one of the virtual design communities on the internet through the tool of web crawler, and stored in a relational database as structuring data.Computer programming languages such as SQL are used to process the structuring data and calculate the statistics.XY scatter plot for characterization function and year is drawn, and correlation analysis between characterization function and year is carried out to verify the hypotheses. **Conclusion** It pointed out the development advantages and disadvantages of cloud innovation for in recent years, and an open industry design cloud network platform was put forward.

KEY WORDS: industry design; cloud innovation; data mining; virtual design community; hyper-network model; hyper-graph; hypotheses verification; correlation analysis

随着Web2.0技术的深入普及，尤其是近几年来 云计算技术应用面的不断拓宽[1-2]，工业设计领域的

收稿日期： 2014-06-15
基金项目： 国家自然科学基金资助项目(No.51175183)
作者简介： 景广超(1987—)，男，河南林州人，华南理工大学硕士生，主攻云创新设计理论。
通讯作者： 姜立军(1968—)，男，湖南益阳人，博士，华南理工大学教授，主要研究方向为机器视觉、计算机图形学、虚拟现实、云创新理论。

创新模式开始从封闭式设计创新模式向开放式云创新模式进行转变[3]。基于互联网大讨论和共享机制的工业设计云创新，由于出色地集中了大众创新智慧，整合分散创新资源，并兼具有自发性的特点，必将深刻地影响未来工业设计乃至整个设计行业的发展。对互联网虚拟设计社区的云创新趋势作定量化的分析和探索，可以对当前设计行业的发展进行明确定位，并预见性地判断行业未来的走向，具有学术意义和实用价值。

这里提出虚拟设计社区的超网络模型，以虚拟在线设计社区为对象，使用网络爬虫工具收集相关研究数据，并利用计算机编程语言对超网络模型中的相关参量进行统计学分析和计算，可视化输出研究结果，以期对在线设计模式提供有益的参考和建议。

1　虚拟设计社区与云创新

虚拟社区也称为在线社区，最早由英国学者Howard Rheingold于1993年提出[4]。Rheingold将虚拟社区定义为：当足够多的人们带着饱满的情感长期进行公开讨论，以期望在赛博空间中形成个人的关系网时，在网络中所出现的社会集合体[4-5]。这一定义在工业设计领域产生了广泛的影响，虚拟设计社区的形式和功能也在不断地进化演变。

典型的虚拟设计社区主要采取BBS（Bulletin Board System）论坛的方式，此种社区的数据流主要组织结构为线索[6]。典型BBS社区的数据流结构见图1[6]。

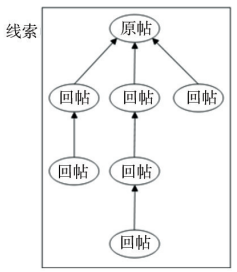

图1　典型BBS社区的数据流结构

Fig.1　Typical data flow structure for BBS community

工业设计BBS中，一个设计主题（以线索的形式出现）如果吸引了众多成员（云创新网络中称之为广义设计师）发散性的广泛参与，以完成某一设计任务或者碰撞出其他全新的设计想法与方案，那么这个设计的产出模式就是一种云创新的模式，但只是初级形态的云创新。随着社会化大生产的全球蔓延，虚拟设计社区技术层面的不断突破，以及工业设计本身的现实需要，不远的将来云创新这种符合时代潮流的大规模创新设计模式会呈现爆炸性的发展。本研究对此提出3项假设并进行定量的统计学检验。假设H1：虚拟设计社区中，设计主题所吸引来参与的广义设计师人数比例近年来呈增长态势；假设H2：虚拟设计社区中，广义设计师所参与的设计主题数量比例近年来呈增长态势；假设H3：虚拟设计社区中，广义设计师对设计主题的相对参与频率近年来呈现显著增长态势。

2　社区网络的超图拓扑结构

2.1　社区网络的超图拓扑结构

虚拟设计社区的网络结构包括大量的多端原件、多层次网络及其功能块，是一种由多节点网络互联而成的超网络[7]。这种超网络可以使用超图来进行拓扑表示。

超图是离散数学中用于研究有限集合中各元素之间的多元关系的一个概念，最初是由法国数学家Berge在1973年提出的。数学上超图的严格定义为[8]：设$X=\{x_1,x_2,\cdots,x_n\}$是一个有限集，若$e_i\neq(i=1,2,\cdots,m)$，$\bigcup_{i=1}^{m} e_i = X$，则称二元关系$H=(E,X)$为一个超图。其中：有限集$X$中的元素$x_1,x_2,\cdots x_n$称为超图的顶点，亦即超网络中的节点；$E=\{e_1,e_2,\cdots,e_m\}$是超图的超边集合；集合$e_i=\{x_{i1},x_{i2},\cdots,x_{ij}\}(i=1,2,\cdots,m)$则称为超图的超边。在超图中，以超边表示的网络连接不限于两个节点之间的简单直线连接，而是可以包含多个节点的多元连接关系。虚拟设计社区的超图拓扑（部分）见图2。

在虚拟设计社区网络中，令$D=\{d_1,d_2,\cdots,d_n\}$表示n位广义设计师所构成的集合，而$T=\{t_1,t_2,\cdots,t_m\}$则表示社区网络中的m个设计主题所构成的集合。那么，在这个由n位广义设计师和m个设计主题所组成的超网络中，其中的一个超图$H=(T,D)$，亦即$H=(\{t_1,t_2,\cdots,t_m\},\{d_1,d_2,\cdots,d_n\})$为以广义设计师为网络节点的超图。

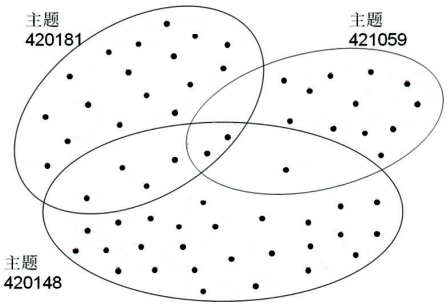

图2 虚拟设计社区的超图拓扑(部分)

Fig.2 Hyper-graph topology for virtual design community(part)

2.2 社区网络的矩阵语义表达

虚拟设计社区的拓扑超图为一个有着n个顶点、m条边的超图,若边j与顶点i关联,a_{ij}=1,否则a_{ij}=0,则由元素a_{ij}(i=1,2,…,n;j=1,2,…,m)构成的$n\cdot m$阶矩阵为超图的关联矩阵,下文记为$\boldsymbol{A}(\boldsymbol{H})$。虚拟设计社区中活跃存在着18 634名广义设计师,11 540个设计主题,那么就构成了一个18 634×11 540的关联矩阵$\boldsymbol{A}(\boldsymbol{H})$。可以用下面的函数$\varphi$来表征社区中设计主题平均所吸引来参与的广义设计师人数比例:

$$\varphi = \frac{1}{m \cdot n}\sum_{j=1}^{m}\sum_{i=1}^{n} a_{ij} \tag{1}$$

同样,用函数ψ来表征社区中广义设计师人均所参与的设计主题数量比例:

$$\psi = \frac{1}{n \cdot m}\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij} \tag{2}$$

3 互联网数据挖掘

这里选取了一个业界知名的工业设计论坛的“设计讨论”板块作为研究的数据样本提取源,使用网络爬虫对该板块内共涉及的18 634名用户,11 540个设计主题的相关详细数据进行了采集。

3.1 社区数据采集与处理

本次数据采集的实验时间是2013年4月4日到2013年4月6日,所收集的对象数据发表时间为2001年1月1日00:00:00到2012年12月31日23:59:59。所收集的数据对象包括:在上述时间段内在设计讨论板块中活动的所有用户的相关信息、该时间段内在该板块中发表的所有主题信息、该时间段内在该板块中发表的所有帖子信息。

3.2 数据模型与数据库存储设计

互联网络上存在的论坛网站是一种复杂的数据源,并且具有其自身的组织特点,可以看作是一种庞大的、混沌的、异构的分布式数据库[9]。收集的论坛数据结构化的数据模型见图3。

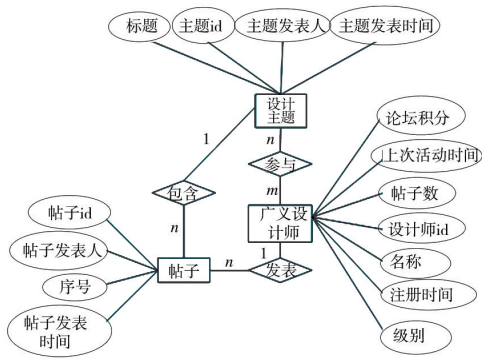

图3 虚拟设计社区的E-R数据模型

Fig.3 E-R data model of virtual design community

4 工业设计云创新现状与趋势分析

4.1 社区发展表征函数定量计算

从采集实验中所获取的数据中取2008~2012年这5年的网络数据,对每年的函数φ和函数ψ进行计算,得到的结果见表1。

表1 虚拟设计社区表征函数值

Tab.1 Characteristic function data for virtual design community

年份	表征函数 φ	表征函数 ψ	表征函数 f(次·月$^{-1}$·人$^{-1}$)
2008	0.003 264	0.030 69	0.000 534 6
2009	0.002 808	0.018 95	0.000 802 6
2010	0.002 539	0.026 41	0.000 924 5
2011	0.002 696	0.039 16	0.001 425 3
2012	0.003 372	0.065 55	0.001 822 9

为使结果最大程度地接近真实情况,在计算过程中历年所取的广义设计师人数n,并非注册人数,而是

当年度在“设计讨论”板块有过活动（如发帖，回帖等）的账号。例如，某账号在2007年某日注册，其最后活动时间为2008年某日，那么在2009年的统计数据中，使用SQL语句：select * from `users` where `reg_date`<'2010-01-01' and `last_date`>'2008-12-31'对数据库进行查询，则对2009年及以后各年份的分析就不再包含此用户信息。

为了考察虚拟设计社区的主题参与频率，需要将采集到的论坛回帖时间数据转换为统一的Unix时间戳进行计算（Unix Timestamp，下文记为t_u，单位为s）。采集对象为回复数不小于3的主题。为了把论坛自身的经营状况（会员数量）等因素对研究结果的影响最大程度地降低，参与频率的统计量应当考虑广义设计师人数的因素，如下：

$$f = \frac{1}{\Delta t_u \cdot n} \tag{3}$$

其中Δt_u为Unix时间戳表示的帖子平均回复时间间隔，n为当年度活跃的广义设计师人数，函数f的单位为“次/（秒·人）”。为使结果更加直观形象，将其单位等价变换为“次/（月·人）”，结果如表1。

φ函数、ψ函数、f函数等3个函数值作为统计量，可以分别对H1假设、H2假设、H3假设进行统计验证。3个函数和年份的xy散点图见图4。

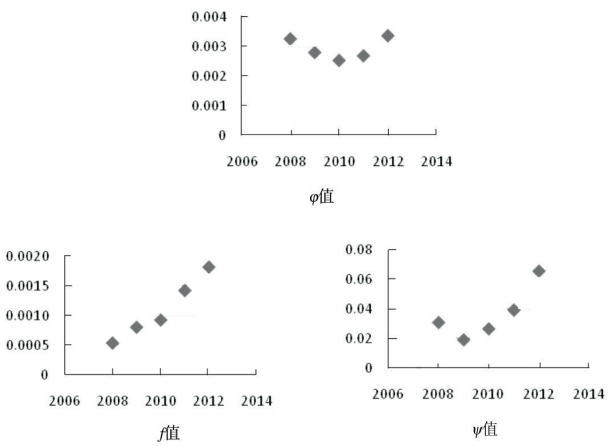

图4 函数统计量对年份的xy散点图

Fig.4 XY Scatter diagram for functional statistics and years

4.2 统计量的假设检验

从图4的散点图分布情况来看，函数ψ、函数f这两个函数与年份之间呈现较为明显的正相关性，而函数φ与时间的相关性并不明显，因此使用SPSS软件对函数统计量与年份作双变量的双侧检验相关分析，得到相关分析结果，见表2。

表2 函数统计量与年份的相关分析

Tab.2 Correlation analysis for functional statistics and years

	Year	函数φ	函数ψ	函数f
Pearson相关性	1	0.044	0.791	0.979
显著性（双侧）		0.943	0.111	0.004
样本数	5	5	5	5

结果表明：（1）函数φ与年份呈现弱正相关，Pearson相关系数为0.044，表明虽然相关程度较弱，但是仍然能够支持假设H1的观点，亦即设计主题所吸引来参与的广义设计师人数比例近年来呈增长态势，可以认为假设H1成立；（2）函数ψ与年份呈现明显正相关，Pearson相关系数为0.791，表明广义设计师所参与的设计主题数量比例近年来呈显著增长态势（p=0.111），假设H2成立；（3）函数f与年份呈现强正相关，Pearson相关系数为0.979，表明近年来，虚拟设计社区中主题的参与频率显著提升（p=0.004），广义设计师越来越及时而快速地参与到设计主题的云端设计，假设H3成立。

5 结语

通过分析可以看出，最近几年来，基于互联网的在线设计社区中，由于自发性地聚集了大批的专业设计人才和设计爱好人士，成员参与到设计主题的大讨论与云设计的意愿日趋强烈。与此同时，互联网设计主题对设计人才的吸纳能力也趋于上升。虚拟设计社区中的大讨论与云设计的强度大幅度提升，成员所建立起来的设计主题能够更加快速地得到其他成员的响应，从而产生出云创新所大量需要地设计交互。这种交互是一种社会网络所隐含的大量知识及虚拟组织参与的知识进化的过程，而这些知识往往隐含了突破性的创新理念，甚至会在螺旋上升的进化过程中获得最佳解答。在Web2.0持续快速发展的今天，无论对于企业和个人，工业设计云创新这一设计模式，将会成为极具竞争力的创新模式和极有价值的创新手

段。尽管如此，研究过程中也暴露出了云创新这一新生事物目前所存在的不足之处和发展瓶颈。首先，互联网虚拟设计社区形式过于单一。这种社区主体目前仍然使用专业性不强、功能性单调的网上论坛的模式，成员主要通过图片、文字这两种方式进行设计探讨，很大程度上限制了成员的积极性和创造性。专业化的云创新网络社区工具呼之欲出，这也对设计研究者提出了一个新的挑战。其次，虚拟设计社区的人员成分复杂，设计环境不够单纯。这与当前社区的成员过滤机制和淘汰机制有关。再次，虚拟设计社区中成员设计水平良莠不齐，但却没有完善的评价系统，这在一定程度上也挫伤了成员的积极性和归属感。如何建立起良性的社区激励机制，这也成为一个亟待解决的课题。

鉴于此，开发一种开放式创新云网络平台，以构建更为专业化的在线虚拟工业设计社区，建立良好的创新云生态，将不仅是有意义的，而且迫在眉睫[10]。基于系统论思想[11]和Web 2.0技术，以产品语义学[12]作为概念表达主方式的平台，可以构建基于用户间协作互助关系的知识创新服务系统，这种环境有助于激发用户的交流兴趣和学习能力，促进用户自服务的发展，为用户的创新思维形成提供更为广阔的空间，也为实现由社会网络参与工业产品创新设计提供了美好前景。

参考文献：

[1] LEE Y C, ZOMAYA A Y.Energy Efficient Utilization of Resources in Cloud Computing Systems[J].The Journal of Supercomputing, 2012, 60(2): 268—280.

[2] XU Xun.From Cloud Computing to Cloud Manufacturing[J]. Robotics and Computer-Integrated Manufacturing, 2012, 28(1): 75—86.

[3] 任丽梅，黄斌.云创新——21世纪的创新模式[M].北京：中共中央党校出版社，2010.
REN Li-mei, HUANG Bin.Cloud Innovation: Model of Innovation in the 21st Century[M].Beijing: The Central Party School Publishing House, 2010.

[4] 王陆.虚拟学习社区的社会网络结构[M].北京：北京大学出版社，2011.
WANG Lu.Social Network Structure of the Virtual Learning Community[M].Beijing: Beijing University Press, 2011.

[5] RHEINGOLD H.The Virtual Community: Homesteading on the Electronic Frontier[M].Addison-Wesley, 1993.

[6] 郭志刚，席耀一，李弼程，等.论坛数据形式化表示技术研究[J].信息工程大学学报，2011，12(6)：734—744.
GUO Zhi-gang, XI Yao-yi, LI Bi-cheng, et al.Research on Forum Data Formalization Expression Technology[J].Journal of the PLA Information Engineering University, 2011, 12(6): 734—744.

[7] 尚艳超，王恒山，王艳灵.基于微博上信息传播的超网络模型[J].技术与创新管理，2012，33(2)：175—178.
SHANG Yan-chao, WANG Heng-shan, WANG Yan-ling. The Supernetwork Model of Information Transmission Based on the Microblog[J].Technology and Innovation Management, 2012, 33(2): 175—178.

[8] 王众托.关于超网络的一点思考[J].上海理工大学学报，2011，33(3)：229—237.
WANG Zhong-tuo.Reflection on Supernetwork[J].Journal of University of Shanghai for Science and Technology, 2011, 33(3): 229—237.

[9] 张芷源.基于超网络的互联网社会网络分析模型的研究[D].上海：上海交通大学，2012.
ZHANG Zhi-yuan.Research on Supernetwork-based Internet Social Network Analysis Model[D].Shanghai: Shanghai Jiaotong University, 2012.

[10] LEE S M, HWANG T, CHOI D. Open Innovation in the Public Sector of Leading Countries[J].Management Decision, 2012, 50(1): 147—162.

[11] 刘秀云.基于系统论语境下的工业设计[J].包装工程，2012，33(16)：85—87.
LIU Xiu-yun.Industrial Design Based on the Context of System Theory[J].Packaging Engineering, 2012, 33(16): 85—87.

[12] 刘胜志，朱钟炎.产品语义学和产品设计[J].包装工程，2006，27(1)：182—195.
LIU Sheng-zhi, ZHU Zhong-yan.Product Semantics and Product Design[J].Packaging Engineering, 2006, 27(1): 182—195.